

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

Voice Based Information Retrieval system

Kiruthika M¹, Priyadarsini S², Rishwana Roshan K³, Shifana Parvin V.M⁴, Dr.G.Umamaheswari⁵

U.G. Student, Department of Electronics & Communication Engineering, PSGCT, Tamil Nadu, India^{1,2,3,4}

Assistant Professor, Department of Electronics & Communication Engineering, PSGCT, Tamil Nadu, India⁵

ABSTRACT:Information retrieval is the process of getting data from database and updating the knowledge with current affairs. This is intended for visually challenged people who are dependent on others to get information. A speech recognizer is an intricate device that responds to human speech. The main objective is to develop a handy device that helps the visually challenged people to get information in others absence. In this paper, the technique of responding to speech input with appropriate answers is presented. The system recognizes the speech input, converts into text, executes the equivalent command and gives the speech output. This is a two-step process. First is speech to text conversion and the second is text to speech conversion. The resultant speech is the answer for question raised.

KEYWORDS:Speech recognition, Text analysis, Hidden Markov Model, MFCC, OCR, Connected component analysis.

I. INTRODUCTION

Speech recognition is a technology that recognizes speech, allowing voice to serve as the main interface between the human and computer. There is a need for a system to help the visually challenged people to provide with information like date, time, weather condition, current happenings and other general information like first search data on the internet. Speech recognition is commonly used to execute commands or write without using a keyboard, mouse or buttons. In real world many tools have been introduced to address visually challenged people's problems using computer vision. But providing the same through a simple mobile with internet is the actual need. There are many types of voice recognition systems in use like user dependant, independent, discrete speech recognition, continuous speech recognition, natural language recognition etc. Out of this, recognising Natural language is difficult because recognition is done instantaneously and no other process like pattern matching is performed to identify the sounds. Such kind of system is developed that works 24 hours a day and 365 days a year. As input and output are in terms of speech, proposed system is user-friendly for anyone who is busy with other tasks, such as driving, cooking etc.

II. LITERATURE REVIEW

A number of speech recognition techniques such as linear timescaled word template matching, dynamic time warped word-template matching (find the phonemes, assemble into words and assemble into sentences) and hidden Markov models (HMM) were used in olden times. Of all the available techniques, HMMs are currently yielding the best performance due to its computational speed and accuracy. The HMM based on MFCC coefficient computation is performed on the input signal. The recorded signal is compared with the original signal where the method of MFCC is used in feature extraction [1] and the result stored is saved as text document. This type of pattern matching is adopted only in few cases where the questions are pre-entered. Recognition of Natural language is performed by Frame blocking, Window blocking using FFT algorithm [2]. Recognition of scanned documents using OCR is now generally considered to be a solved problem in identifying the wordings in some scripts. Components of an OCR system consist of optical scanning, binarization, segmentation, feature extraction and recognition, connected component analysis [3]. This enables to transform the text to speech and give the speech output. Android based speech recognition is used now a days in mobile which uses Google API's for the conversion [4]. It explains the pseudo code, used to access the Google server to get the information from the repositories. This method is adopted in the later part to fetch information from the internet once the speech is recognised using HMM.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

III. PROPOSED METHODOLOGY

a) Speech to Text conversion

Sometimes the voice recorded may not be recognised. Noise and other environmental factors may disturb the operation. So the convenient way of recognising is done using HMM model which extracts the features like pitch, tone, frequency etc. from the sound and converts to text using MFCC (Mel-Frequency Cepstrum Coefficients) that gives information in the spectral form delivering a simple and distinctive illustration of the spectral properties of the voice signal which. After the speech characteristics has been changed into MFCC matrices features are extracted, segmented and then subsequently the words are recognized and articulated out.

HMMs are used in speech recognition because a speech signal can be viewed as a piecewise stationary signal or a short-time stationary signal. In a short time-scale like 10 milliseconds, speech can be approximated as a stationary process. Speech can be thought of as a Markov model for many stochastic purposes. HMMs are popular because they can be trained automatically and are simple and computationally feasible to use. Here HMM would output a sequence of n-dimensional real-valued vectors, outputting one of these every 10 milliseconds. The vectors would consist of cepstral coefficients, which are obtained by taking a Fourier transform of a short time window of speech and de-correlating the spectrum using a cosine transform, then taking the most significant coefficients. The hidden Markov model will tend to have in each state a statistical distribution that is a mixture of diagonal covariance Gaussians, which will give a likelihood for each observed vector. Each word or each phoneme, will have a different output distribution; a hidden Markov model for a sequence of words or phonemes is made by concatenating the individual trained hidden Markov models for the separate words and phonemes. Likewise, each separate words are recognised and saved as text.

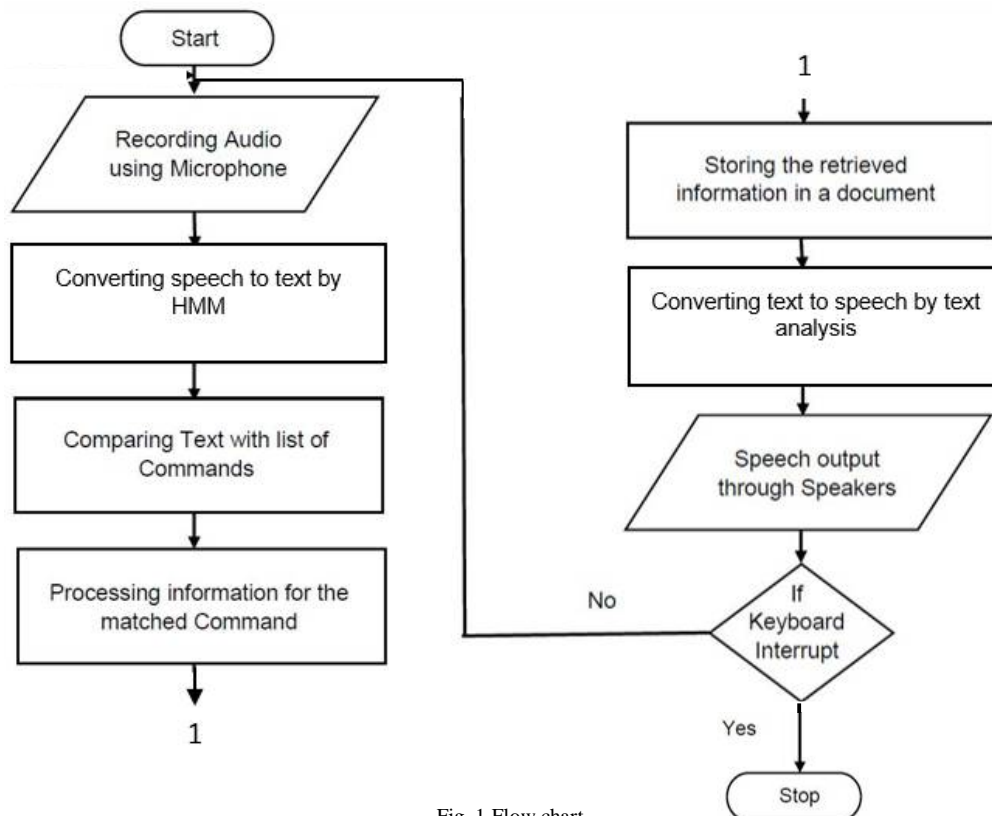


Fig. 1 Flow chart

b) Text to Speech conversion

Now resultant text is stored in a word document. Next the content in the document is compared with the set of commands stored in the database which has an access to the code that is currently executed. Each code has its own

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

functionality like finding the date, time, weather etc from the processor. For basic questions regarding historic data, current affairs, the code is directed to browse the internet and find the answers for the same that is given as output in the form of speech. The steps involved in the text to speech conversion are as follows.

1. OCR: OCR is used to extract the features and recognize the text. The pre-processing step involves binarisation in which binary image of the text is obtained and then features are extracted via canny edge detection and Harris corner algorithm. Once the text is identified it is subjected to phonetic conversion.

2. Phonetic conversion: Phonetics is a predefined library which has a set of sounds for each word or letter that is given as input. Once the words from the documents match with the syllable the sound is given as output. During continuous speech, the articulator movements depend on the preceding and the following phonemes. Thus by this process the text is converted to speech by matching with phonetics.

3. Prosodic Phrasing: Once the sound for each syllable is obtained the words are phrased. In most languages the written text does not correspond to its pronunciation, so that in order to describe correct pronunciation some kind of symbolic presentation is needed. Every language has a different phonetic alphabet and a different set of possible phonemes and their combinations.

4. Concatenate Speech: The sounds of the text are concatenated to give a meaningful output as speech. It does this by connected component analysis and then the output is observed at the speakers. By the end of this process text is converted to speech as portrayed in Fig. 1.

IV. SYSTEM ARCHITECTURE

The system mainly includes Raspberry Pi 3 B+ because of its functionalities like Wi-Fi, audio output, image processing capabilities etc. Input is given through Microphone and output is taken via Speaker as shown in Fig 2. To view the screen, HDMI to VGA Converter is connected between Raspberry Pi and monitor. Raspberry Pi is powered using 5V power supply.

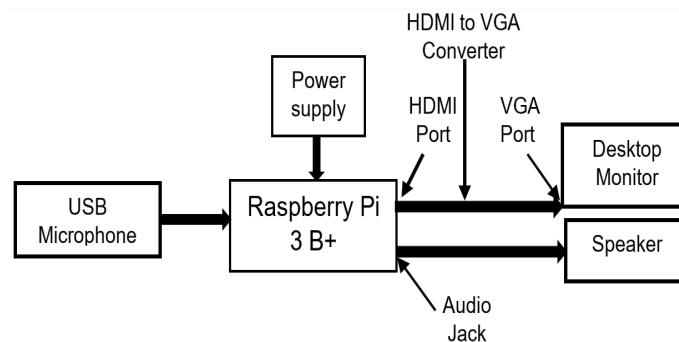


Fig. 2 Block Diagram for Voice based information retrieval system.

OpenCV is a library of programming functions mainly aimed at real-time computer vision which is written in C++. There are bindings in Python, Java and MATLAB. The API for these interfaces can be found in the online documentation. Python language is preferred because it allows execution without any delay and has python wrappers which are used as python modules. 'os' and 'sys' modules provide numerous tools to deal with filenames, paths, directories. 'cv2' module is a Python interface. New Out Of Box Software (NOOBS) is an easy operating system installation manager for the Raspberry Pi to install OS like Raspbian, Pidora etc.

V. EXPERIMENTAL RESULTS

There are two modes of operation based on the way of executing the code.

1. Execution in Standalone system.
2. Execution via mobile phone.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

Overall setup of the system is made as displayed in Fig. 3 which follows operational mode 1. Raspberry pican also be SSH'ed with laptop using LXTerminal (Xserver) or can be interfaced with desktop monitor using VGA port as shown in Fig. 2. Python code is executed on the terminal that's displayed in monitor. This mode is highly used for performing more complex computation like solving algebraic equations, sending mail, chatting in social media etc. as it allows interfaces like Keyboard, mouse.

In order to add the portability feature to it, operational mode 2 is followed where Raspberry pi's terminal is obtained in Mobile phone with "Juice SSH" app installed in it, as shown in Fig. 4. Internet connection is mandatory in both the mode of operation.



Fig. 3 Execution in Standalone system

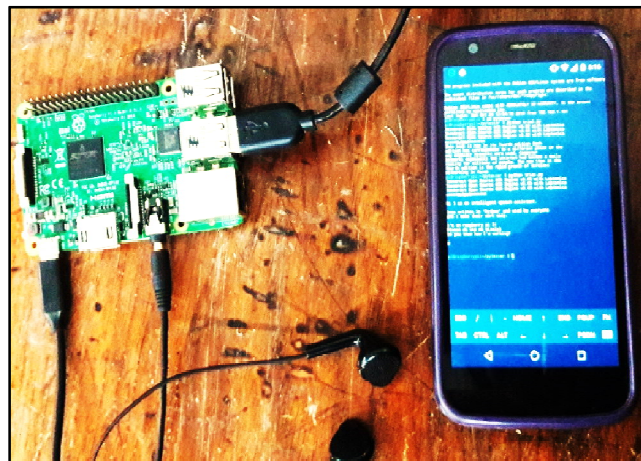


Fig. 4 Execution via mobile phone.

Each time the user has to wait for the sound "Say something" to query his question. Once recognised, the device replies with "Got it! Now to recognize it..". Then the answer for queried question is read out as shown in Fig. 5. It takes about 2.42 seconds to answer a question. The questions can be pre-entered for the answers to help the user to better know about other information that Google can't answer. It can be user name, his relations, address, mobile number and other related personal information as shown in Fig. 5a).

```
kiruthika@kiruthika-Vostro-3558:~$ python w.py
A moment of silence, please...
Set minimum energy threshold to 467.027812912
Say something!
Got it! Now to recognize it...
You said good morning
good morning have a nice day
Say something!
Got it! Now to recognize it...
You said how are you
Ia fine dude How are you
```

a)

```
Say something!
Got it! Now to recognize it...
You said what is the temperature at Coimbatore
27 deg C heat index: 30 deg C
26 minutes ago
```

b)

```
Say something!
Got it! Now to recognize it...
You said what is the time
8:53:24 pm IST | Saturday, October 14, 2017
```

c)

```
Say something!
Got it! Now to recognize it...
You said what is the capital of India
New Delhi, Delhi, India
```

d)

```
Say something!
Got it! Now to recognize it...
You said 8 x 80
640
Say something!
Got it! Now to recognize it...
You said 4 - 2
2
```

e)

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 12, December 2017

```
Say something!
Got it! Now to recognize it...
You said define machine
In abstract topology, a machine is method for producing infinite loop spaces and
an abstract machine that is implemented in hardware is simply called a machine.
Say something!
Got it! Now to recognize it...
You said what is the meaning of Raspberry Pi
brand | Raspberry Pi
model | A
available | since Monday, February 4, 2013
price | $25 (US dollars)
CPU | ARM1176JZF-S (ARMv6k) 700 MHz (megahertz)
memory | 256 MB (megabytes)
graphics | Broadcom VideoCore IV
power | 1.5 W (watts)
Say something!
Got it! Now to recognize it...
You said who is the president of India
Ram Nath Kovind (from 25/07/2017 to present)
Say something!
Got it! Now to recognize it...
You said who is the Prime Minister of India
Narendra Modi (from 26/05/2014 to present)
```

f)

Fig. 5 Output window with answers for a) Pre-entered questions b) Weather c) Time d) Historic data e) Calculation f) Current affairs and basic questions

It is not necessary to hold the entire device all-time. It can be made simpler by having Bluetooth Microphone with the user, to send the input to the device which is about 20-30 mt distance and the output is taken via Bluetooth speaker which can be placed near the user.

VI. CONCLUSION & FUTURE SCOPE

Voice based information retrieval system is an asset for visually challenged people and its portability feature helps them in a better way. This system successfully answers all types of question and the same is verified even under noisy environment. As a future work, the system can be secured with face recognition to prevent from opening and viewing personal accounts of others in Facebook, Twitter etc. Also it can be further developed to guide visually challenged people to walk independently on roads with the help of sensors and maps. This type of devices can be fitted in reception area to deliver information for the customers in the absence of receptionist in the office.

REFERENCES

1. RupaliS Chavan, Ganesh. S Sable, "An Overview of Speech Recognition Using HMM", IJCSMC, Vol. 2, Issue. 6, June 2013.
2. Mr. Sanjay Bhardwaj, Mr. Sunil Pathania , Mr. Rajesh Akela, "Speech Recognition using Hidden Markov Model and Viterbi Algorithm", IJARECE, ISSN:2278 – 909X ,Vol 4, Issue 5, May 2015.
3. JishaGopinath, Aravind S, "Text to Speech Conversion System using OCR", IJETAE, ISSN 2250-2459, Vol 5, Issue 1, Jan 2015.
4. Lavin Jalan, Rahul Masrani, RoshanJadhav and TejaswiniPalav, "Speech Recognition based Learning System", International Journal of Engineering Trends and Technology(IJETT), Vol. 4, No. 2, Dec 2013.
5. DarpanRaut, SumedhSalvi, ShrikrishnaSalvi and Prof. S. R. Rangari, "Intelligent Voice Assistant using Androidplatform", International Journal of Computer Trends andTechnology (IJCTT) , Vol. 34, No. 3, Apr 2016.
6. SutarShekhar S, KamadNeha S and DeokateLaxmanJ, "Intelligent Voice Assistant Using Android Platform", International Journal of Advanced Research in Computer Science and Management Studies (IJARCSMS), Vol. 3, No. 3, Mar 2015.
7. AditiBhalerao, Samira Bhilare, AnaghaBondade and MonalShingade, "Smart Voice Assistant: a universalvoice control solution for non-visual access to the Android operating system", International ResearchJournal of Engineering and Technology (IRJET), Vol. 4, Jan 2017.